(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/16635

(22) International Filing Date: 23 May 2002 (23.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/294,076   29 May 2001 (29.05.2001)   US

(71) Applicant: LEXICON GENETICS INCORPORATED [US/US]; 4000 Research Forest Drive, The Woodlands, TX 77381 (US).

(72) Inventors: YU, Xuanchuan; 6 Danville Crossing, Conroe, TX 77385 (US). MIRANDA, Maricar; 14814 Belterraza Drive, Houston, TX 77083 (US). HU, Yi; 333 Holly Creek Ct. #203, The Woodlands, TX 77381 (US).

(74) Agents: ISHIMOTO, Lance, K. et al.; Lexicon Genetics Incorporated, 4000 Research Forest Drive, The Woodlands, TX 77381 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL HUMAN HYDROXYLASES AND POLYNUCLEOTIDES ENCODING THE SAME

(57) Abstract: Novel human polynucleotide and polypeptide sequences are disclosed that can be used in industrial, therapeutic, diagnostic, and pharmacogenomic applications.

NOVEL HUMAN HYDROXYLASES AND
POLYNUCLEOTIDES ENCODING THE SAME

The present application claims the benefit of U.S.
5 Provisional Application Number 60/294,076, which was
filed on May 29, 2001, and is herein incorporated by
reference in its entirety.

1. INTRODUCTION

10 The present invention relates to the discovery,
identification, and characterization of novel human
polynucleotides encoding proteins sharing sequence
similarity with mammalian hydroxylases. The invention
encompasses the described polynucleotides, host cell
15 expression systems, the encoded protein, fusion
proteins, polypeptides and peptides, antibodies to the
encoded proteins and peptides, and genetically
engineered animals that either lack or overexpress the
disclosed polynucleotides, antagonists and agonists of
20 the proteins, and other compounds that modulate the
expression or activity of the proteins encoded by the
disclosed polynucleotides, which can be used for
diagnosis, drug screening, clinical trial monitoring,
the treatment of diseases and disorders, and cosmetic or
25 nutriceutical applications.

2. BACKGROUND OF THE INVENTION

Hydroxylases are enzymes that mediate the cleavage
of hydroxyl groups, and many hydroxylases act in
30 biological pathways suitable for drug intervention.

3. SUMMARY OF THE INVENTION

The present invention relates to the discovery,
identification, and characterization of nucleotides that
35 encode novel human proteins, and the corresponding amino
acid sequences of these proteins. The novel human
proteins (NHPs) described for the first time herein

1

share structural similarity with animal hydroxylases, and particularly tryptophan hydroxylases, which are involved in a rate-limiting step in the biosynthesis of a number of neurologically active compounds, including,

5 but not limited to, DOPA, serotonin and melatonin. Given the well known physiological functions of the above compounds, the described hydroxylases can be used to identify and/or develop agents useful for modulating behavior (i.e., treating anxiety, depression,

10 hyperactivity, sleep disorders, etc.). The novel human nucleic acid (cDNA) sequences described herein encode proteins/open reading frames (ORFs) of 490, 486, 485, and 484 amino acids in length (SEQ ID NOS:2, 4, 6, and 8, respectively).

15   In addition to the direct therapeutic use of the described NHPs, the invention also encompasses agonists and antagonists of the described NHPs, including small molecules, large molecules, mutant NHPs, or portions thereof, that compete with native NHPs, peptides, and

20 antibodies, as well as nucleotide sequences that can be used to inhibit the expression of the described NHPs (e.g., antisense and ribozyme molecules, and open reading frame or regulatory sequence replacement constructs) or to enhance the expression of the

25 described NHPs (e.g., expression constructs that place the described polynucleotide under the control of a strong promoter system), and transgenic animals that express a NHP sequence, or "knock-outs" (which can be conditional) that do not express a functional NHP.

30 Knock-out mice can be produced in several ways, one of which involves the use of mouse embryonic stem cell ("ES cell") lines that contain gene trap mutations in a murine homolog of at least one of the described NHPs. When the unique NHP sequences described in SEQ ID NOS:1-

35 8 are "knocked-out" they provide a method of identifying phenotypic expression of the particular gene, as well as

2

a method of assigning function to previously unknown
genes. In addition, animals in which the unique NHP
sequences described in SEQ ID NOS:1-8 are "knocked-out"
provide a unique source in which to elicit antibodies to
5   homologous and orthologous proteins, which would have
been previously viewed by the immune system as "self"
and therefore would have failed to elicit significant
antibody responses.

     Additionally, the unique NHP sequences described in
10 SEQ ID NOS:1-8 are useful for the identification of
protein coding sequences, and mapping a unique gene to a
particular chromosome. These sequences identify
biologically verified exon splice junctions, as opposed
to splice junctions that may have been bioinformatically
15 predicted from genomic sequence alone. The sequences of
the present invention are also useful as additional DNA
markers for restriction fragment length polymorphism
(RFLP) analysis, and in forensic biology, particularly
given the presence of a nucleotide polymorphism within
20 the described sequences.

     Further, the present invention also relates to
processes for identifying compounds that modulate, *i.e.*,
act as agonists or antagonists of, NHP expression and/or
NHP activity that utilize purified preparations of the
25 described NHPs and/or NHP products, or cells expressing
the same. Such compounds can be used as therapeutic
agents for the treatment of any of a wide variety of
symptoms associated with biological disorders or
imbalances.

30

     4.   DESCRIPTION OF THE SEQUENCE LISTING AND FIGURES
     The Sequence Listing provides the sequences of the
ORFs encoding the described NHP amino acid sequences.

5. DETAILED DESCRIPTION OF THE INVENTION

The NHPs described for the first time herein are
novel proteins that can be expressed in, *inter alia*,
5   human cell lines, human fetal brain, spinal cord, lymph
node, prostate, testis, thyroid, pancreas, pericardium,
hypothalamus, fetal kidney, fetal lung, 6 and 9-week old
embryos, osteosarcoma, and embryonic carcinoma cells.

The present invention encompasses the nucleotides
10 presented in the Sequence Listing, host cells expressing
such nucleotides, the expression products of such
nucleotides, and: (a) nucleotides that encode mammalian
homologs of the described polynucleotides, including the
specifically described NHPs, and the NHP products;
15 (b) nucleotides that encode one or more portions of a
NHP that correspond to functional domains of a NHP, and
the polypeptide products specified by such nucleotide
sequences, including, but not limited to, the novel
regions of any active domain(s); (c) isolated
20 nucleotides that encode mutant versions, engineered or
naturally occurring, of a described NHP, in which all or
a part of at least one domain is deleted or altered, and
the polypeptide products specified by such nucleotide
sequences, including, but not limited to, soluble
25 proteins and peptides; (d) nucleotides that encode
chimeric fusion proteins containing all or a portion of
a coding region of a NHP, or one of its domains (*e.g.*, a
receptor or ligand binding domain, accessory
protein/self-association domain, *etc.*) fused to another
30 peptide or polypeptide; or (e) therapeutic or diagnostic
derivatives of the described polynucleotides, such as
oligonucleotides, antisense polynucleotides, ribozymes,
dsRNA, or gene therapy constructs, comprising a sequence
first disclosed in the Sequence Listing.

35   As discussed above, the present invention
includes the human DNA sequences presented in the

4

Sequence Listing (and vectors comprising the same), and
additionally contemplates any nucleotide sequence
encoding a contiguous NHP open reading frame (ORF) that
hybridizes to a complement of a DNA sequence presented

5 in the Sequence Listing under highly stringent
conditions, e.g., hybridization to filter-bound DNA in
0.5 M NaHPO$_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA
at 65°C, and washing in 0.1x SSC/0.1% SDS at 68°C
(Ausubel et al., eds., 1989, Current Protocols in

10 Molecular Biology, Vol. I, Green Publishing Associates,
Inc., and John Wiley & Sons, Inc., N.Y., at p. 2.10.3)
and encodes a functionally equivalent expression
product. Additionally contemplated are any nucleotide
sequences that hybridize to the complement of a DNA

15 sequence that encodes and expresses an amino acid
sequence presented in the Sequence Listing under
moderately stringent conditions, e.g., washing in 0.2x
SSC/0.1% SDS at 42°C (Ausubel et al., 1989, supra), yet
still encode a functionally equivalent NHP product.

20 Functional equivalents of a NHP include naturally
occurring NHPs present in other species, and mutant
NHPs, whether naturally occurring or engineered (by site
directed mutagenesis, gene shuffling, directed evolution
as described in, for example, U.S. Patent Nos. 5,837,458

25 and 5,723,323 both of which are herein incorporated by
reference in their entirety). The invention also
includes degenerate nucleic acid variants of the
disclosed NHP polynucleotide sequences.

Additionally contemplated are polynucleotides

30 encoding a NHP ORF, or its functional equivalent,
encoded by a polynucleotide sequence that is about 99,
95, 90, or about 85 percent similar or identical to
corresponding regions of the nucleotide sequences of the
Sequence Listing (as measured by BLAST sequence

35 comparison analysis using, for example, the GCG sequence

5

analysis package, as described herein, using standard default settings).

The invention also includes nucleic acid molecules, preferably DNA molecules, that hybridize to, and are

5 therefore the complements of, the described NHP nucleotide sequences. Such hybridization conditions may be highly stringent or less highly stringent, as described herein. In instances where the nucleic acid molecules are deoxyoligonucleotides ("DNA oligos"), such

10 molecules are generally about 16 to about 100 bases long, or about 20 to about 80 bases long, or about 34 to about 45 bases long, or any variation or combination of sizes represented therein that incorporate a contiguous region of sequence first disclosed in the Sequence

15 Listing. Such oligonucleotides can be used in conjunction with the polymerase chain reaction (PCR) to screen libraries, isolate clones, and prepare cloning and sequencing templates, etc.

Alternatively, such NHP oligonucleotides can be

20 used as hybridization probes for screening libraries, and assessing gene expression patterns (particularly using a microarray or high-throughput "chip" format). Additionally, a series of NHP oligonucleotide sequences, or the complements thereof, can be used to represent all

25 or a portion of the described NHP sequences. An oligonucleotide or polynucleotide sequence first disclosed in at least a portion of one or more of the sequences of SEQ ID NOS:1-8 can be used as a hybridization probe in conjunction with a solid support

30 matrix/substrate (resins, beads, membranes, plastics, polymers, metal or metallized substrates, crystalline or polycrystalline substrates, etc.). Of particular note are spatially addressable arrays (i.e., gene chips, microtiter plates, etc.) of oligonucleotides and

35 polynucleotides, or corresponding oligopeptides and polypeptides, wherein at least one of the biopolymers

6

present on the spatially addressable array comprises an
oligonucleotide or polynucleotide sequence first
disclosed in at least one of the sequences of SEQ ID
NOS:1-8, or an amino acid sequence encoded thereby.

5   Methods for attaching biopolymers to, or synthesizing
biopolymers on, solid support matrices, and conducting
binding studies thereon, are disclosed in, *inter alia*,
U.S. Patent Nos. 5,700,637, 5,556,752, 5,744,305,
4,631,211, 5,445,934, 5,252,743, 4,713,326, 5,424,186,

10   and 4,689,405, the disclosures of which are herein
incorporated by reference in their entirety.

Addressable arrays comprising sequences first
disclosed in SEQ ID NOS:1-8 can be used to identify and
characterize the temporal and tissue specific expression

15   of a gene. These addressable arrays incorporate
oligonucleotide sequences of sufficient length to confer
the required specificity, yet be within the limitations
of the production technology. The length of these
probes is usually within a range of between about 8 to

20   about 2000 nucleotides. Preferably the probes consist
of 60 nucleotides, and more preferably 25 nucleotides,
from the sequences first disclosed in SEQ ID NOS:1-8.

For example, a series of such NHP oligonucleotide
sequences, or the complements thereof, can be used in

25   chip format to represent all or a portion of the
described sequences. The oligonucleotides, typically
between about 16 to about 40 (or any whole number within
the stated range) nucleotides in length, can partially
overlap each other, and/or the sequence may be

30   represented using oligonucleotides that do not overlap.
Accordingly, the described polynucleotide sequences
shall typically comprise at least about two or three
distinct oligonucleotide sequences of at least about 8
nucleotides in length that are each first disclosed in

35   the described Sequence Listing. Such oligonucleotide
sequences can begin at any nucleotide present within a

7

sequence in the Sequence Listing, and proceed in either a sense (5'-to-3') orientation vis-a-vis the described sequence or in an antisense orientation.

Microarray-based analysis allows the discovery of
5 broad patterns of genetic activity, providing new understanding of gene functions, and generating novel and unexpected insight into transcriptional processes and biological mechanisms. The use of addressable arrays comprising sequences first disclosed in SEQ ID
10 NOS:1-8 provides detailed information about transcriptional changes involved in a specific pathway, potentially leading to the identification of novel components, or gene functions that manifest themselves as novel phenotypes.

15 Probes consisting of sequences first disclosed in SEQ ID NOS:1-8 can also be used in the identification, selection, and validation of novel molecular targets for drug discovery. The use of these unique sequences permits the direct confirmation of drug targets, and
20 recognition of drug dependent changes in gene expression that are modulated through pathways distinct from the intended target of the drug. These unique sequences therefore also have utility in defining and monitoring both drug action and toxicity.

25 As an example of utility, the sequences first disclosed in SEQ ID NOS:1-8 can be utilized in microarrays, or other assay formats, to screen collections of genetic material from patients who have a particular medical condition. These investigations can
30 also be carried out using the sequences first disclosed in SEQ ID NOS:1-8 *in silico*, and by comparing previously collected genetic databases and the disclosed sequences using computer software known to those in the art.

Thus the sequences first disclosed in SEQ ID NOS:1-
35 8 can be used to identify mutations associated with a

8

particular disease, and also in diagnostic or prognostic assays.

Although the presently described sequences have been specifically described using nucleotide sequence,

5   it should be appreciated that each of the sequences can uniquely be described using any of a wide variety of additional structural attributes, or combinations thereof. For example, a given sequence can be described by the net composition of the nucleotides present within

10   a given region of the sequence, in conjunction with the presence of one or more specific oligonucleotide sequence(s) first disclosed in SEQ ID NOS:1-8. Alternatively, a restriction map specifying the relative positions of restriction endonuclease digestion sites,

15   or various palindromic or other specific oligonucleotide sequences, can be used to structurally describe a given sequence. Such restriction maps, which are typically generated by widely available computer programs (*e.g.,* the University of Wisconsin GCG sequence analysis

20   package, SEQUENCHER 3.0, Gene Codes Corp., Ann Arbor, MI, *etc.*), can optionally be used in conjunction with one or more discrete nucleotide sequence(s) present in the sequence that can be described by the relative position of the sequence relative to one or more

25   additional sequence(s) or one or more restriction sites present in the disclosed sequence.

For oligonucleotide probes, highly stringent conditions may refer, *e.g.,* to washing in 6x SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligos), 48°C

30   (for 17-base oligos), 55°C (for 20-base oligos), and 60°C (for 23-base oligos). These nucleic acid molecules may encode or act as NHP antisense molecules, useful, for example, in NHP gene regulation and/or as antisense primers in amplification reactions of NHP nucleic acid

35   sequences. With respect to NHP gene regulation, such techniques can be used to regulate biological functions.

9

Further, such sequences may be used as part of ribozyme
and/or triple helix sequences that are also useful for
NHP gene regulation.

Inhibitory antisense or double stranded

5 oligonucleotides can additionally comprise at least one
modified base moiety that is selected from the group
including, but not limited to, 5-fluorouracil,
5-bromouracil, 5-chlorouracil, 5-iodouracil,
hypoxanthine, xanthine, 4-acetylcytosine,

10 5-(carboxyhydroxylmethyl) uracil,
5-carboxymethylaminomethyl-2-thiouridine,
5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-
galactosylqueosine, inosine, N6-isopentenyladenine,
1-methylguanine, 1-methylinosine, 2,2-dimethylguanine,

15 2-methyladenine, 2-methylguanine, 3-methylcytosine,
5-methylcytosine, N6-adenine, 7-methylguanine,
5-methylaminomethyluracil, 5-methoxyaminomethyl-
2-thiouracil, beta-D-mannosylqueosine,
5'-methoxycarboxymethyluracil, 5-methoxyuracil,

20 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic
acid (v), wybutoxosine, pseudouracil, queosine,
2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil,
4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid
methylester, uracil-5-oxyacetic acid (v), 5-methyl-

25 2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil,
(acp3)w, and 2,6-diaminopurine.

The antisense oligonucleotide can also comprise at
least one modified sugar moiety selected from the group
including, but not limited to, arabinose,

30 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the antisense
oligonucleotide will comprise at least one modified
phosphate backbone selected from the group including,
but not limited to, a phosphorothioate, a

35 phosphorodithioate, a phosphoramidothioate, a
phosphoramidate, a phosphordiamidate, a

10

methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

In yet another embodiment, the antisense oligonucleotide is an α-anomeric oligonucleotide. An α-
5 anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gautier et al., 1987, Nucl. Acids Res. 15:6625-6641). The oligonucleotide is a 2'-0-methylribonucleotide (Inoue
10 et al., 1987, Nucl. Acids Res. 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue et al., 1987, FEBS Lett. 215:327-330). Alternatively, double stranded RNA can be used to disrupt the expression and function of a targeted NHP.

15 Oligonucleotides of the invention can be synthesized by standard methods known in the art, e.g., by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.). As examples, phosphorothioate
20 oligonucleotides can be synthesized by the method of Stein et al. (1988, Nucl. Acids Res. 16:3209), and methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin et al., 1988, Proc. Natl. Acad. Sci. USA 85:7448-7451),
25 etc.

Low stringency conditions are well-known to those of skill in the art, and will vary predictably depending on the specific organisms from which the library and the labeled sequences are derived. For guidance regarding
30 such conditions see, for example, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (and periodic updates thereof), and Ausubel et al., 1989, supra.

Alternatively, suitably labeled NHP nucleotide
35 probes can be used to screen a human genomic library using appropriately stringent conditions or by PCR. The

11

identification and characterization of human genomic
clones is helpful for identifying polymorphisms
(including, but not limited to, nucleotide repeats,
microsatellite alleles, single nucleotide polymorphisms,
5   or coding single nucleotide polymorphisms), determining
the genomic structure of a given locus/allele, and
designing diagnostic tests. For example, sequences
derived from regions adjacent to the intron/exon
boundaries of the human gene can be used to design
10 primers for use in amplification assays to detect
mutations within the exons, introns, splice sites (e.g.,
splice acceptor and/or donor sites), etc., that can be
used in diagnostics and pharmacogenomics.

   For example, the present sequences can be used in
15 restriction fragment length polymorphism (RFLP) analysis
to identify specific individuals. In this technique, an
individual's genomic DNA is digested with one or more
restriction enzymes, and probed on a Southern blot to
yield unique bands for identification (as generally
20 described in U.S. Patent No. 5,272,057, incorporated
herein by reference). In addition, the sequences of the
present invention can be used to provide polynucleotide
reagents, e.g., PCR primers, targeted to specific loci
in the human genome, which can enhance the reliability
25 of DNA-based forensic identifications by, for example,
providing another "identification marker" (i.e., another
DNA sequence that is unique to a particular individual).
Actual base sequence information can be used for
identification as an accurate alternative to patterns
30 formed by restriction enzyme generated fragments.

   Further, a NHP homolog can be isolated from nucleic
acid from an organism of interest by performing PCR
using two degenerate or "wobble" oligonucleotide primer
pools designed on the basis of amino acid sequences
35 within the NHP products disclosed herein. The template
for the reaction may be genomic DNA, or total RNA, mRNA,

12

and/or cDNA obtained by reverse transcription of mRNA prepared from human or non-human cell lines or tissue known to express, or suspected of expressing, an allele of a NHP gene.

5   The PCR product can be subcloned and sequenced to ensure that the amplified sequences represent the sequence of the desired NHP gene. The PCR fragment can then be used to isolate a full length cDNA clone by a variety of methods. For example, the amplified fragment

10   can be labeled and used to screen a cDNA library, such as a bacteriophage cDNA library. Alternatively, the labeled fragment can be used to isolate genomic clones via the screening of a genomic library.

        PCR technology can also be used to isolate full

15   length cDNA sequences. For example, RNA can be isolated, following standard procedures, from an appropriate cellular or tissue source (*i.e.*, one known to express, or suspected of expressing, a NHP gene, such as, for example, testis tissue). A reverse

20   transcription (RT) reaction can be performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis. The resulting RNA/DNA hybrid may then be "tailed" using a standard terminal transferase

25   reaction, the hybrid may be digested with RNase H, and second strand synthesis may then be primed with a complementary primer. Thus, cDNA sequences upstream of the amplified fragment can be isolated. For a review of cloning strategies that can be used, see, *e.g.*, Sambrook

30   *et al.*, 1989, *supra*.

        A cDNA encoding a mutant NHP sequence can be isolated, for example, by using PCR. In this case, the first cDNA strand may be synthesized by hybridizing an oligo-dT oligonucleotide to mRNA isolated from tissue

35   known to express, or suspected of expressing, a NHP, in an individual putatively carrying a mutant NHP allele,

13

and by extending the new strand with reverse
transcriptase. The second strand of the cDNA is then
synthesized using an oligonucleotide that hybridizes
specifically to the 5' end of the normal sequence.

5   Using these two primers, the product is then amplified
via PCR, optionally cloned into a suitable vector, and
subjected to DNA sequence analysis through methods well-
known to those of skill in the art. By comparing the
DNA sequence of the mutant NHP allele to that of a

10   corresponding normal NHP allele, the mutation(s)
responsible for the loss or alteration of function of
the mutant NHP gene product can be ascertained.

Alternatively, a genomic library can be constructed
using DNA obtained from an individual suspected of

15   carrying, or known to carry, a mutant NHP allele (e.g.,
a person manifesting a NHP-associated phenotype such as,
for example, neurological disorders, such as, but not
limited to, depression, anxiety, Alzheimer's disease or
Parkinson's disease, obesity, high blood pressure,

20   connective tissue disorders, infertility, developmental
disorders, etc.), or a cDNA library can be constructed
using RNA from a tissue known to express, or suspected
of expressing, a mutant NHP allele. A normal NHP gene,
or any suitable fragment thereof, can then be labeled

25   and used as a probe to identify the corresponding mutant
NHP allele in such libraries. Clones containing mutant
NHP sequences can then be purified and subjected to
sequence analysis according to methods well-known to
those skilled in the art.

30   Additionally, an expression library can be
constructed utilizing cDNA synthesized from, for
example, RNA isolated from a tissue known to express, or
suspected of expressing, a mutant NHP allele in an
individual suspected of carrying, or known to carry,

35   such a mutant allele. In this manner, gene products
made by the putatively mutant tissue can be expressed

14

and screened using standard antibody screening
techniques in conjunction with antibodies raised against
a normal NHP product, as described below (for screening
techniques, see, for example, Harlow and Lane, eds.,
5   1988, "Antibodies: A Laboratory Manual", Cold Spring
Harbor Press, Cold Spring Harbor, N.Y.).

Additionally, screening can be accomplished by
screening with labeled NHP fusion proteins, such as, for
example, alkaline phosphatase-NHP or NHP-alkaline
10   phosphatase fusion proteins. In cases where a NHP
mutation results in an expression product with altered
function (e.g., as a result of a missense or a
frameshift mutation), polyclonal antibodies to a NHP are
likely to cross-react with a corresponding mutant NHP
15   expression product. Library clones detected via their
reaction with such labeled antibodies can be purified
and subjected to sequence analysis according to methods
well-known in the art.

The invention also encompasses: (a) DNA vectors
20   that contain any of the foregoing NHP coding sequences
and/or their complements (i.e., antisense); (b) DNA
expression vectors that contain any of the foregoing NHP
coding sequences operatively associated with a
regulatory element that directs the expression of the
25   coding sequences (for example, baculovirus as described
in U.S. Patent No. 5,869,336 herein incorporated by
reference); (c) genetically engineered host cells that
contain any of the foregoing NHP coding sequences
operatively associated with a regulatory element that
30   directs the expression of the coding sequences in the
host cell; and (d) genetically engineered host cells
that express an endogenous NHP sequence under the
control of an exogenously introduced regulatory element
(i.e., gene activation). As used herein, regulatory
35   elements include, but are not limited to, inducible and
non-inducible promoters, enhancers, operators, and other

15

elements known to those skilled in the art that drive and regulate expression. Such regulatory elements include, but are not limited to, the cytomegalovirus (hCMV) immediate early gene, regulatable, viral elements

5 (particularly retroviral LTR promoters), the early or late promoters of SV40 or adenovirus, the *lac* system, the *trp* system, the *TAC* system, the *TRC* system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, the promoter for

10 3-phosphoglycerate kinase (PGK), the promoters of acid phosphatase, and the promoters of the yeast α-mating factors.

The present invention also encompasses antibodies and anti-idiotypic antibodies (including Fab fragments),

15 antagonists and agonists of a NHP, as well as compounds or nucleotide constructs that inhibit expression of a NHP sequence (transcription factor inhibitors, antisense and ribozyme molecules, or open reading frame sequence or regulatory sequence replacement constructs), or

20 promote the expression of a NHP (*e.g.*, expression constructs in which NHP coding sequences are operatively associated with expression control elements such as promoters, promoter/enhancers, *etc.*).

The NHPs or NHP peptides, NHP fusion proteins, NHP

25 nucleotide sequences, antibodies, antagonists and agonists can be useful for the detection of mutant NHPs, or inappropriately expressed NHPs, for the diagnosis of disease. The NHP proteins or peptides, NHP fusion proteins, NHP nucleotide sequences, host cell expression

30 systems, antibodies, antagonists, agonists and genetically engineered cells and animals can be used for screening for drugs (or high throughput screening of combinatorial libraries) effective in the treatment of the symptomatic or phenotypic manifestations of

35 perturbing the normal function of a NHP in the body. The use of engineered host cells and/or animals may

16

offer an advantage in that such systems allow not only
for the identification of compounds that bind to the
endogenous receptor for a NHP, but can also identify
compounds that trigger NHP-mediated activities or
5   pathways.

        Finally, the NHP products can be used as
therapeutics. For example, soluble derivatives such as
NHP peptides/domains corresponding to a NHP, NHP fusion
protein products (especially NHP-Ig fusion proteins,
10 i.e., fusions of a NHP, or a domain of a NHP, to an
IgFc), NHP antibodies and anti-idiotypic antibodies
(including Fab fragments), antagonists or agonists
(including compounds that modulate or act on downstream
targets in a NHP-mediated pathway), can be used to
15 directly treat diseases or disorders. For instance, the
administration of an effective amount of a soluble NHP
(or, for example, a suitably derivatized NHP, e.g., with
polyethylene glycol (PEG), albumin, etc.), a NHP-IgFc
fusion protein, or an anti-idiotypic antibody (or its
20 Fab) that mimics the NHP, could activate or effectively
antagonize the endogenous NHP receptor. Nucleotide
constructs encoding such NHP products can be used to
genetically engineer host cells to express such products
in vivo; these genetically engineered cells function as
25 "bioreactors" in the body delivering a continuous supply
of a NHP, a NHP peptide, or a NHP fusion protein to the
body. Nucleotide constructs encoding functional NHPs,
mutant NHPs, as well as antisense and ribozyme
molecules, can also be used in "gene therapy" approaches
30 for the modulation of NHP expression. Thus, the
invention also encompasses pharmaceutical formulations
and methods for treating biological disorders.

        Various aspects of the invention are described in
greater detail in the subsections below.
35

17

5.1 THE NHP SEQUENCES

The cDNA sequences and the corresponding deduced amino acid sequences of the described NHPs are presented in the Sequence Listing. The described sequences were

5 compiled from human gene trapped sequences, and cDNAs made from human fetal brain, fetal lung, and lymph node mRNAs (Edge Biosystems, Gaithersburg, MD). As biologically validated (spliced and polyadenylated sequences), the described sequences are useful for

10 mapping the corresponding coding region of the human genome and particularly for mapping exon splice junctions. The described NHPs are apparently encoded on human chromosome 12 (see GENBANK accession no. AC023966).

15   An A/T polymorphism was detected during the sequencing of the NHPs at the nucleotide position represented by, for example, position 1125 of SEQ ID NO:1 (and the corresponding location in the other NHP nucleotide sequences), both of which result in an ala at

20 the region corresponding to amino acid (aa) position 375 of, for example, SEQ ID NO:2 (and the corresponding location in the other NHP amino acid sequences). As these polymorphisms are coding single nucleotide polymorphisms (SNPs), they are particularly useful in

25 forensic analysis.

An additional application of the described novel human polynucleotide sequences is their use in the molecular mutagenesis/evolution of proteins that are at least partially encoded by the described novel sequences

30 using, for example, polynucleotide shuffling or related methodologies. Such approaches are described in U.S. Patent Nos. 5,830,721 and 5,837,458, which are herein incorporated by reference in their entirety.

NHP gene products can also be expressed in

35 transgenic animals. Animals of any species, including, but not limited to, worms, mice, rats, rabbits, guinea

18

pigs, pigs, micro-pigs, birds, goats, and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees, may be used to generate NHP transgenic animals.

Any technique known in the art may be used to
5 introduce a NHP transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Hoppe and Wagner, 1989, U.S. Patent No. 4,873,191); retrovirus-mediated gene transfer into germ
10 lines (Van der Putten *et al.*, 1985, Proc. Natl. Acad. Sci. USA *82*:6148-6152); gene targeting in embryonic stem cells (Thompson *et al.*, 1989, Cell *56*:313-321); electroporation of embryos (Lo, 1983, Mol Cell. Biol. *3*:1803-1814); and sperm-mediated gene transfer
15 (Lavitrano *et al.*, 1989, Cell *57*:717-723); *etc.* For a review of such techniques, see Gordon, 1989, Transgenic Animals, Intl. Rev. Cytol. *115*:171-229, which is incorporated by reference herein in its entirety.

The present invention provides for transgenic
20 animals that carry a NHP transgene in all their cells, as well as animals that carry a transgene in some, but not all their cells, *i.e.*, mosaic animals or somatic cell transgenic animals. A transgene may be integrated as a single transgene, or in concatamers, *e.g.*, head-to-
25 head tandems or head-to-tail tandems. A transgene may also be selectively introduced into and activated in a particular cell-type by following, for example, the teaching of Lasko *et al.*, 1992, Proc. Natl. Acad. Sci. USA *89*:6232-6236. The regulatory sequences required for
30 such a cell-type specific activation will depend upon the particular cell-type of interest, and will be apparent to those of skill in the art.

When it is desired that a NHP transgene be integrated into the chromosomal site of the endogenous
35 NHP gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing

19

some nucleotide sequences homologous to the endogenous NHP gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide

5 sequence of the endogenous NHP gene (*i.e.*, "knockout" animals).

The transgene can also be selectively introduced into a particular cell-type, thus inactivating the endogenous NHP gene in only that cell-type, by

10 following, for example, the teaching of Gu *et al.*, 1994, Science 265:103-106. The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell-type of interest, and will be apparent to those of skill in the art.

15 Once transgenic animals have been generated, the expression of the recombinant NHP gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to assay whether integration

20 of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques that include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in*

25 *situ* hybridization analysis, and RT-PCR. Samples of NHP gene-expressing tissue may also be evaluated immunocytochemically using antibodies specific for the NHP transgene product.

The present invention also provides for "knock-in"

30 animals. Knock-in animals are those in which a polynucleotide sequence (*i.e.*, a gene or a cDNA) that the animal does not naturally have in its genome is inserted in such a way that it is expressed. Examples include, but are not limited to, a human gene or cDNA

35 used to replace its murine ortholog in the mouse, a murine cDNA used to replace the murine gene in the

20

mouse, and a human gene or cDNA or murine cDNA that is
tagged with a reporter construct used to replace the
murine ortholog or gene in the mouse. Such replacements
can occur at the locus of the murine ortholog or gene,

5   or at another specific site. Such knock-in animals are
useful for the *in vivo* study, testing and validation of,
intra alia, human drug targets, as well as for compounds
that are directed at the same, and therapeutic proteins.

10   5.2 NHPS AND NHP POLYPEPTIDES
         NHPs, NHP polypeptides, NHP peptide fragments,
mutated, truncated, or deleted forms of the NHPs, and/or
NHP fusion proteins can be prepared for a variety of
uses. These uses include, but are not limited to, the

15 generation of antibodies, as therapeutic products
(particularly in the treatment of behavioral disorders
such as, but not limited to, dementia, insomnia,
depression, anorexia, and premenstrual syndrome, *etc.*,
as well as cancers such as, but not limited to,

20 lymphoma, leukemia, sarcomas, carcinomas, myelomas,
etc.), as reagents in diagnostic assays, for the
identification of other cellular gene products related
to a NHP, as reagents in assays for screening for
compounds that can be used as pharmaceutical reagents

25 useful in the therapeutic treatment of mental,
biological, or medical disorders and disease. Given the
similarity information and expression data, the
described NHPs can be targeted (by drugs, oligos,
antibodies, *etc.*) in order to treat disease, or to

30 therapeutically augment the efficacy of therapeutic or
chemotherapeutic agents. In addition, the described
NHPs can be used in drug screening assays similar to
those described in, for example, U.S. Patent No.
6,048,850, herein incorporated by reference, in order to

35 identify compounds for treating diseases such as, for

21

example, depression, anxiety, immune disorders,
Alzheimer's disease, epilepsy, and Parkinson's disease.

The Sequence Listing discloses the amino acid
sequences encoded by the described NHP polynucleotides.
5 The NHPs display initiator methionines in DNA sequence
contexts consistent with translation initiation sites,
and a N-terminal signal sequence characteristic of
secreted or membrane proteins (the signal sequence is
often cleaved from the mature form of the protein
10 incident to the protein translocating across the
membrane).

The NHP amino acid sequences of the invention
include the amino acid sequences presented in the
Sequence Listing, as well as analogues and derivatives
15 thereof, as well as any oligopeptide sequence of at
least about 10-40 amino acids, about 12-35 amino acids,
or about 16-30 amino acids in length first disclosed in
the Sequence Listing. Further, corresponding NHP
homologues from other species are encompassed by the
20 invention. In fact, any NHP encoded by the NHP
nucleotide sequences described herein are within the
scope of the invention, as are any novel polynucleotide
sequences encoding all or any novel portion of an amino
acid sequence presented in the Sequence Listing. The
25 degenerate nature of the genetic code is well-known,
and, accordingly, each amino acid presented in the
Sequence Listing is generically representative of the
well-known nucleic acid "triplet" codon, or in many
cases codons, that can encode the amino acid. As such,
30 as contemplated herein, the amino acid sequences
presented in the Sequence Listing, when taken together
with the genetic code (see, for example, Table 4-1 at
page 109 of "Molecular Cell Biology", 1986, J. Darnell
et al., eds., Scientific American Books, New York, N.Y.,
35 herein incorporated by reference), are generically
representative of all the various permutations and

22

combinations of nucleic acid sequences that can encode such amino acid sequences.

The invention also encompasses proteins that are functionally equivalent to the NHPs encoded by the
5   presently described nucleotide sequences, as judged by any of a number of criteria, including, but not limited to, the ability to bind and cleave a substrate of a NHP, the ability to effect an identical or complementary downstream pathway, or a change in cellular metabolism
10 (e.g., proteolytic activity, ion flux, tyrosine phosphorylation, etc.). Such functionally equivalent NHP proteins include, but are not limited to, additions or substitutions of amino acid residues within the amino acid sequence encoded by the NHP nucleotide sequences
15 described herein, but that result in a silent change, thus producing a functionally equivalent expression product. Amino acid substitutions can be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic
20 nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine,
25 asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

A variety of host-expression vector systems can be
30 used to express the NHP nucleotide sequences of the invention. Where, as in the present instance, the NHP products or NHP polypeptides can be produced in soluble or secreted forms (by removing one or more transmembrane domains where applicable), the peptide or polypeptide
35 can be recovered from the culture media. Such expression systems also encompass engineered host cells

23

that express a NHP, or a functional equivalent, *in situ*.
Purification or enrichment of a NHP from such expression
systems can be accomplished using appropriate detergents
and lipid micelles and methods well-known to those

5 skilled in the art. However, such engineered host cells
themselves may be used in situations where it is
important not only to retain the structural and
functional characteristics of a NHP, but to assess
biological activity, *e.g.*, in certain drug screening

10 assays.

The expression systems that may be used for
purposes of the invention include, but are not limited
to, microorganisms such as bacteria (*e.g.*, *E. coli*, *B.
subtilis*) transformed with recombinant bacteriophage

15 DNA, plasmid DNA or cosmid DNA expression vectors
containing NHP nucleotide sequences; yeast (*e.g.*,
Saccharomyces, *Pichia*) transformed with recombinant
yeast expression vectors containing NHP nucleotide
sequences; insect cell systems infected with recombinant

20 virus expression vectors (*e.g.*, baculovirus) containing
NHP nucleotide sequences; plant cell systems infected
with recombinant virus expression vectors (*e.g.*,
cauliflower mosaic virus, CaMV; tobacco mosaic virus,
TMV) or transformed with recombinant plasmid expression

25 vectors (*e.g.*, Ti plasmid) containing NHP nucleotide
sequences; or mammalian cell systems (*e.g.*, COS, CHO,
BHK, 293, 3T3) harboring recombinant expression
constructs containing NHP nucleotide sequences and
promoters derived from the genome of mammalian cells

30 (*e.g.*, metallothionein promoter) or from mammalian
viruses (*e.g.*, the adenovirus late promoter; the
vaccinia virus 7.5K promoter).

In bacterial systems, a number of expression
vectors may be advantageously selected depending upon

35 the use intended for the NHP product being expressed.
For example, when a large quantity of such a protein is

24

to be produced for the generation of pharmaceutical
compositions of or containing a NHP, or for raising
antibodies to a NHP, vectors that direct the expression
of high levels of fusion protein products that are
5   readily purified may be desirable. Such vectors
include, but are not limited to, the *E. coli* expression
vector pUR278 (Ruther *et al.*, 1983, EMBO J. *2*:1791), in
which a NHP coding sequence may be ligated individually
into the vector in-frame with the *lacZ* coding region so
10 that a fusion protein is produced; pIN vectors (Inouye
and Inouye, 1985, Nucleic Acids Res. *13*:3101-3109; Van
Heeke and Schuster, 1989, J. Biol. Chem. *264*:5503-5509);
and the like. pGEX vectors (Pharmacia or American Type
Culture Collection) can also be used to express foreign
15 polypeptides as fusion proteins with glutathione
S-transferase (GST). In general, such fusion proteins
are soluble and can easily be purified from lysed cells
by adsorption to glutathione-agarose beads followed by
elution in the presence of free glutathione. The pGEX
20 vectors are designed to include thrombin or factor Xa
protease cleavage sites so that the cloned target
expression product can be released from the GST moiety.

In an exemplary insect system, *Autographa
californica* nuclear polyhedrosis virus (AcNPV) is used
25 as a vector to express foreign polynucleotide sequences.
The virus grows in *Spodoptera frugiperda* cells. A NHP
coding sequence can be cloned individually into a non-
essential region (for example the polyhedrin gene) of
the virus and placed under control of.an AcNPV promoter
30 (for example the polyhedrin promoter). Successful
insertion of a NHP coding sequence will result in
inactivation of the polyhedrin gene and production of
non-occluded recombinant virus (*i.e.*, virus lacking the
proteinaceous coat coded for by the polyhedrin gene).
35 These recombinant viruses are then used to infect
Spodoptera frugiperda cells in which the inserted

25

sequence is expressed (*e.g.*, see Smith *et al.*, 1983, J.
Virol. *46*:584; Smith, U.S. Patent No. 4,215,051).

In mammalian host cells, a number of viral-based
expression systems may be utilized. In cases where an

5 adenovirus is used as an expression vector, the NHP
nucleotide sequence of interest may be ligated to an
adenovirus transcription/translation control complex,
e.g., the late promoter and tripartite leader sequence.
This chimeric sequence may then be inserted in the

10 adenovirus genome by *in vitro* or *in vivo* recombination.
Insertion in a non-essential region of the viral genome
(*e.g.*, region E1 or E3) will result in a recombinant
virus that is viable and capable of expressing a NHP
product in infected hosts (*e.g.*, see Logan and Shenk,

15 1984, Proc. Natl. Acad. Sci. USA *81*:3655-3659).
Specific initiation signals may also be required for
efficient translation of inserted NHP nucleotide
sequences. These signals include the ATG initiation
codon and adjacent sequences. In cases where an entire

20 NHP gene or cDNA, including its own initiation codon and
adjacent sequences, is inserted into the appropriate
expression vector, no additional translational control
signals may be needed. However, in cases where only a
portion of a NHP coding sequence is inserted, exogenous

25 translational control signals, including, perhaps, the
ATG initiation codon, may be provided. Furthermore, the
initiation codon should be in phase with the reading
frame of the desired coding sequence to ensure
translation of the entire insert. These exogenous

30 translational control signals and initiation codons can
be of a variety of origins, both natural and synthetic.
The efficiency of expression may be enhanced by the
inclusion of appropriate transcription enhancer
elements, transcription terminators, *etc.* (see Bitter *et*

35 *al.*, 1987, Methods in Enzymol. *153*:516-544).

26

In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the expression product in the specific fashion desired. Such modifications (e.g.,

5   glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins and expression

10 products. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed. To this end, eukaryotic host cells that possess the cellular machinery for the desired processing of the primary

15 transcript, glycosylation, and phosphorylation of the expression product may be used. Such mammalian host cells include, but are not limited to, CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, and in particular, human cell lines.

20 For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines that stably express the NHP sequences described herein can be engineered. Rather than using expression vectors that contain viral origins of

25 replication, host cells can be transformed with DNA controlled by appropriate expression control elements (e.g., promoter, enhancer sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the

30 foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into

35 their chromosomes and grow to form foci, which in turn can be cloned and expanded into cell lines. This method

27

may advantageously be used to engineer cell lines that
express a NHP product. Such engineered cell lines may
be particularly useful in screening and evaluation of
compounds that affect the endogenous activity of a NHP
5 product.

A number of selection systems may be used,
including, but not limited to, the herpes simplex virus
thymidine kinase (Wigler et al., 1977, Cell 11:223),
hypoxanthine-guanine phosphoribosyltransferase
10 (Szybalska and Szybalski, 1962, Proc. Natl. Acad. Sci.
USA 48:2026), and adenine phosphoribosyltransferase
(Lowy et al., 1980, Cell 22:817) genes, which can be
employed in tk⁻, hgprt⁻ or aprt⁻ cells, respectively.
Also, antimetabolite resistance can be used as the basis
15 of selection for the following genes: dhfr, which
confers resistance to methotrexate (Wigler et al., 1980,
Proc. Natl. Acad. Sci. USA 77:3567; O'Hare et al., 1981,
Proc. Natl. Acad. Sci. USA 78:1527); gpt, which confers
resistance to mycophenolic acid (Mulligan and Berg,
20 1981, Proc. Natl. Acad. Sci. USA 78:2072); neo, which
confers resistance to the aminoglycoside G-418
(Colberre-Garapin et al., 1981, J. Mol. Biol. 150:1);
and hygro, which confers resistance to hygromycin
(Santerre et al., 1984, Gene 30:147).
25 Alternatively, any fusion protein can be readily
purified by utilizing an antibody specific for the
fusion protein being expressed. Another exemplary
system allows for the ready purification of non-
denatured fusion proteins expressed in human cell lines
30 (Janknecht et al., 1991, Proc. Natl. Acad. Sci. USA
88:8972-8976). In this system, the sequence of interest
is subcloned into a vaccinia recombination plasmid such
that the sequence's open reading frame is
translationally fused to an amino-terminal tag
35 consisting of six histidine residues. Extracts from
cells infected with recombinant vaccinia virus are

28

loaded onto Ni²⁺·nitriloacetic acid-agarose columns, and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

Also encompassed by the present invention are
5   fusion proteins that direct a NHP to a target organ and/or facilitate transport across the membrane into the cytosol. Conjugation of NHPs to antibody molecules or their Fab fragments could be used to target cells bearing a particular epitope. Attaching an appropriate
10 signal sequence to a NHP would also transport a NHP to a desired location within the cell. Alternatively targeting of a NHP or its nucleic acid sequence might be achieved using liposome or lipid complex based delivery systems. Such technologies are described in "Liposomes:
15 A Practical Approach", New, R.R.C., ed., Oxford University Press, N.Y., and in U.S. Patent Nos. 4,594,595, 5,459,127, 5,948,767 and 6,110,490 and their respective disclosures, which are herein incorporated by reference in their entirety. Additionally embodied are
20 novel protein constructs engineered in such a way that they facilitate transport of NHPs to a target site or desired organ, where they cross the cell membrane and/or the nucleus where the NHPs can exert their functional activity. This goal may be achieved by coupling of a
25 NHP to a cytokine or other ligand that provides targeting specificity, and/or to a protein transducing domain (see generally U.S. Provisional Patent Application Ser. Nos. 60/111,701 and 60/056,713, both of which are herein incorporated by reference, for examples
30 of such transducing sequences), to facilitate passage across cellular membranes, and can optionally be engineered to include nuclear localization signals.

Additionally contemplated are oligopeptides that are modeled on an amino acid sequence first described in
35 the Sequence Listing. Such NHP oligopeptides are generally between about 10 to about 100 amino acids

long, or between about 16 to about 80 amino acids long,
or between about 20 to about 35 amino acids long, or any
variation or combination of sizes represented therein
that incorporate a contiguous region of sequence first
5   disclosed in the Sequence Listing. Such NHP
oligopeptides can be of any length disclosed within the
above ranges and can initiate at any amino acid position
represented in the Sequence Listing.
        The invention also contemplates "substantially
10 isolated" or "substantially pure" proteins or
polypeptides. By a "substantially isolated" or
"substantially pure" protein or polypeptide is meant a
protein or polypeptide that has been separated from at
least some of those components that naturally accompany
15 it. Typically, the protein or polypeptide is
substantially isolated or pure when it is at least 60%,
by weight, free from the proteins and other naturally-
occurring organic molecules with which it is naturally
associated *in vivo*. Preferably, the purity of the
20 preparation is at least 75%, more preferably at least
90%, and most preferably at least 99%, by weight. A
substantially isolated or pure protein or polypeptide
may be obtained, for example, by extraction from a
natural source, by expression of a recombinant nucleic
25 acid encoding the protein or polypeptide, or by
chemically synthesizing the protein or polypeptide.
        Purity can be measured by any appropriate method,
e.g., column chromatography such as immunoaffinity
chromatography using an antibody specific for the
30 protein or polypeptide, polyacrylamide gel
electrophoresis, or HPLC analysis. A protein or
polypeptide is substantially free of naturally
associated components when it is separated from at least
some of those contaminants that accompany it in its
35 natural state. Thus, a polypeptide that is chemically
synthesized or produced in a cellular system different

30

from the cell from which it naturally originates will be, by definition, substantially free from its naturally associated components. Accordingly, substantially isolated or pure proteins or polypeptides include

5 eukaryotic proteins synthesized in *E. coli*, other prokaryotes, or any other organism in which they do not naturally occur.

5.3 ANTIBODIES TO NHP PRODUCTS

10 Antibodies that specifically recognize one or more epitopes of a NHP, epitopes of conserved variants of a NHP, or peptide fragments of a NHP, are also encompassed by the invention. Such antibodies include, but are not limited to, polyclonal antibodies, monoclonal antibodies

15 (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above.

20 The antibodies of the invention may be used, for example, in the detection of a NHP in a biological sample and may, therefore, be utilized as part of a diagnostic or prognostic technique whereby patients may be tested for abnormal amounts of a NHP. Such

25 antibodies may also be utilized in conjunction with, for example, compound screening schemes for the evaluation of the effect of test compounds on expression and/or activity of a NHP expression product. Additionally, such antibodies can be used in conjunction with gene

30 therapy to, for example, evaluate normal and/or engineered NHP-expressing cells prior to their introduction into a patient. Such antibodies may additionally be used in methods for the inhibition of abnormal NHP activity. Thus, such antibodies may be

35 utilized as a part of treatment methods.

31

For the production of antibodies, various host
animals may be immunized by injection with a NHP, a NHP
peptide (*e.g.*, one corresponding to a functional domain
of a NHP), a truncated NHP polypeptide (a NHP in which
5   one or more domains have been deleted), functional
equivalents of a NHP, or mutated variants of a NHP.
Such host animals may include, but are not limited to,
pigs, rabbits, mice, goats, and rats, to name but a few.
Various adjuvants may be used to increase the
10   immunological response, depending on the host species,
including, but not limited to, Freund's adjuvant
(complete and incomplete), mineral salts such as
aluminum hydroxide or aluminum phosphate, chitosan,
surface active substances such as lysolecithin, pluronic
15   polyols, polyanions, peptides, oil emulsions, and
potentially useful human adjuvants such as BCG (bacille
Calmette-Guerin) and *Corynebacterium parvum*.
Alternatively, the immune response could be enhanced by
combination and/or coupling with molecules such as
20   keyhole limpet hemocyanin, tetanus toxoid, diphtheria
toxoid, ovalbumin, cholera toxin, or fragments thereof.
Polyclonal antibodies are heterogeneous populations of
antibody molecules derived from the sera of the
immunized animals.
25   Monoclonal antibodies, which are homogeneous
populations of antibodies to a particular antigen, can
be obtained by any technique that provides for the
production of antibody molecules by continuous cell
lines in culture. These include, but are not limited
30   to, the hybridoma technique of Kohler and Milstein,
(1975, Nature *256*:495-497; and U.S. Patent No.
4,376,110), the human B-cell hybridoma technique (Kosbor
et al., 1983, Immunology Today *4*:72; Cole *et al.*, 1983,
Proc. Natl. Acad. Sci. USA *80*:2026-2030), and the EBV-
35   hybridoma technique (Cole *et al.*, 1985, Monoclonal
Antibodies And Cancer Therapy, Alan R. Liss, Inc., pp.

32

77-96). Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, and IgD, and any subclass thereof. The hybridomas producing the mAbs of this invention may be cultivated *in vitro* or *in vivo*.

5 Production of high titers of mAbs *in vivo* makes this the presently preferred method of production.

In addition, techniques developed for the production of "chimeric antibodies" (Morrison *et al.*, 1984, Proc. Natl. Acad. Sci. USA *81*:6851-6855; Neuberger

10 *et al.*, 1984, Nature, *312*:604-608; Takeda *et al.*, 1985, Nature, *314*:452-454) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric

15 antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. Such technologies are described in U.S. Patent Nos. 6,114,598, 6,075,181

20 and 5,877,397 and their respective disclosures, which are herein incorporated by reference in their entirety. Also encompassed by the present invention is the use of fully humanized monoclonal antibodies, as described in U.S. Patent No. 6,150,584 and respective disclosures,

25 which are herein incorporated by reference in their entirety.

Alternatively, techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778; Bird, 1988, Science *242*:423-426; Huston *et*

30 *al.*, 1988, Proc. Natl. Acad. Sci. USA *85*:5879-5883; and Ward *et al.*, 1989, Nature *341*:544-546) can be adapted to produce single chain antibodies against NHP expression products. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via

35 an amino acid bridge, resulting in a single chain polypeptide.

33

Antibody fragments that recognize specific epitopes
may be generated by known techniques. For example, such
fragments include, but are not limited to: F(ab')$_2$
fragments, which can be produced by pepsin digestion of

5   an antibody molecule; and Fab fragments, which can be
generated by reducing the disulfide bridges of F(ab')$_2$
fragments. Alternatively, Fab expression libraries may
be constructed (Huse et al., 1989, Science, 246:1275-
1281) to allow rapid and easy identification of

10 monoclonal Fab fragments with the desired specificity.

Antibodies to a NHP can, in turn, be utilized to
generate anti-idiotype antibodies that "mimic" a given
NHP, using techniques well-known to those skilled in the
art (see, e.g., Greenspan and Bona, 1993, FASEB J.

15 7:437-444; and Nissinoff, 1991, J. Immunol. 147:2429-
2438). For example, antibodies that bind to a NHP
domain and competitively inhibit the binding of a NHP to
its cognate receptor can be used to generate anti-
idiotypes that "mimic" the NHP and, therefore, bind and

20 activate or neutralize a receptor. Such anti-idiotypic
antibodies, or Fab fragments of such anti-idiotypes, can
be used in therapeutic regimens involving a NHP-mediated
pathway.

Additionally given the high degree of relatedness

25 of mammalian NHPs, NHP knock-out mice (having never seen
a NHP, and thus never been tolerized to a NHP) have a
unique utility, as they can be advantageously applied to
the generation of antibodies against the disclosed
mammalian NHPs (i.e., a NHP will be immunogenic in NHP

30 knock-out animals).

The present invention is not to be limited in scope
by the specific embodiments described herein, which are
intended as single illustrations of individual aspects
of the invention, and functionally equivalent methods

35 and components are within the scope of the invention.
Indeed, various modifications of the invention, in

34

addition to those shown and described herein, will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims. All

5 cited publications, patents, and patent applications are herein incorporated by reference in their entirety.

35

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:7.

5

2. The isolated nucleic acid molecule of claim 1, wherein the molecule comprises the nucleotide sequence of SEQ ID NO:5.

10

3. The isolated nucleic acid molecule of claim 2, wherein the molecule comprises the nucleotide sequence of SEQ ID NO:3.

4. The isolated nucleic acid molecule of

15 claim 3, wherein the molecule comprises the nucleotide sequence of SEQ ID NO:1.

5. An isolated nucleic acid molecule comprising a nucleotide sequence that:

20
(a) encodes the amino acid sequence shown in SEQ ID NO:4; and
(b) hybridizes under highly stringent conditions to the nucleotide sequence of SEQ ID NO:3 or the complement thereof.

25

6. An isolated nucleic acid molecule encoding the amino acid sequence described in SEQ ID NO:8.

30
7. The isolated nucleic acid molecule of claim 6, wherein the molecule encodes the amino acid sequence of SEQ ID NO:6.

8. The isolated nucleic acid molecule of

35 claim 7, wherein the molecule encodes the amino acid sequence of SEQ ID NO:4.

36

9. The isolated nucleic acid molecule of claim 8, wherein the molecule encodes the amino acid sequence of SEQ ID NO:2.

5

10. A substantially isolated protein having the activity of the protein shown in SEQ ID NOS:2, 4, 6, or 8, which is encoded by a nucleotide sequence that hybridizes to SEQ ID NO:1, 3, 5, or 7 under highly stringent conditions.

10

15

20

37

SEQUENCE LISTING

<110> LEXICON GENETICS INCORPORATED

<120> Novel Human Hydroxylases and Polynucleotides Encoding the Same

<130> LEX-0352-PCT

<150> US 60/294,076
<151> 2001-05-29

<160> 8

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 1473
<212> DNA
<213> homo sapiens

<400> 1
atgcagccag caatgatgat gttttccagt aaatactggg cacggagagg gttttccctg 60
gattcagcag tgcccgaaga gcatcagcta cttggcagct caacactaaa taaacctaac 120
tctggcaaaa atgacgacaa aggcaacaag ggaagcagca aacgtgaagc tgctaccgaa 180
agtggcaaga cagcagttgt tttctccttg aagaatgaag ttggtggatt ggtaaaagca 240
ctgaggctct ttcaggaaaa acgtgtcaac atggttcata ttgaatccag gaaatctcgg 300
cgaagaagtt ctgaggttga aatctttgtg gactgtgagt gtgggaaaac agaattcaat 360
gagctcattc agttgctgaa atttcaaacc actattgtga cgctgaatcc tccagagaac 420
atttggacag aggaagaaga gctagaggat gtgccctggt ccctcggaa gatctctgag 480
ttagacaaat gctctcacag agttctcatg tatggttctg agcttgatgc tgaccaccca 540
ggatttaagg acaatgtcta tcgacagaga agaaagtatt ttgtggatgt ggccatgggt 600
tataaatatg gtcagcccat tcccagggtg gagtatactg aagaagaaac taaaacttgg 660
ggtgttgtat tccgggagct ctccaaactc tatcccactc atgcttgccg agagtatttg 720
aaaaacttcc ctctgctgac taaatactgt ggctacagag aggacaatgt gcctcaactc 780
gaagatgtct ccatgtttct gaaagaaagg tctggcttca cggtgaggcc ggtggctgga 840
tacctgagcc cacgagactt tctggcagga ctggcctaca gagtgttcca ctgtacccag 900
tacatccggc atggctcaga tcccctctac accccagaac cagacacatg ccatgaactc 960
ttgggacatg ttccactact tgcggatcct aagtttgctc agttttcaca agaaataggt 1020
ctggcgtctc tgggagcatc agatgaagat gttcagaaac tagccacgtg ctatttcttc 1080
acaatcgagt ttggcctttg caagcaagaa gggcaactgc gggcatatgg agcaggactc 1140
cttttcctcca ttggagaatt aaagcacgcc ccttctgaca aggcatgtgt gaaagccttt 1200
gacccaaaga caacttgctt acaggaatgc cttatcacca ccttccagga agcctacttt 1260
gtttcagaaa gtttttgaaga agccaaagaa aagatgaggg actttgcaaa gtcaattacc 1320
cgtcccttct cagtatactt caatccctac acacagagta ttgaaattct gaaagacacc 1380
agaagtattg aaaatgtggt gcaggacctt cgcagcgact tgaatacagt gtgtgatgct 1440
ttaaacaaaa tgaaccaata tctggggatt tga   1473

<210> 2
<211> 490
<212> PRT
<213> homo sapiens

<400> 2
Met Gln Pro Ala Met Met Met Phe Ser Ser Lys Tyr Trp Ala Arg Arg
1   5   10   15
Gly Phe Ser Leu Asp Ser Ala Val Pro Glu Glu His Gln Leu Leu Gly
            20   25   30
Ser Ser Thr Leu Asn Lys Pro Asn Ser Gly Lys Asn Asp Asp Lys Gly
        35   40   45
Asn Lys Gly Ser Ser Lys Arg Glu Ala Ala Thr Glu Ser Gly Lys Thr
    50   55   60
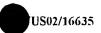Ala Val Val Phe Ser Leu Lys Asn Glu Val Gly Gly Leu Val Lys Ala

```
65                    70                    75                    80
Leu Arg Leu Phe Gln Glu Lys Arg Val Asn Met Val His Ile Glu Ser
                85                    90                    95
Arg Lys Ser Arg Arg Arg Ser Ser Glu Val Glu Ile Phe Val Asp Cys
                100                   105                   110
Glu Cys Gly Lys Thr Glu Phe Asn Glu Leu Ile Gln Leu Leu Lys Phe
                115                   120                   125
Gln Thr Thr Ile Val Thr Leu Asn Pro Pro Glu Asn Ile Trp Thr Glu
    130                   135                   140
Glu Glu Glu Leu Glu Asp Val Pro Trp Phe Pro Arg Lys Ile Ser Glu
145                   150                   155                   160
Leu Asp Lys Cys Ser His Arg Val Leu Met Tyr Gly Ser Glu Leu Asp
                165                   170                   175
Ala Asp His Pro Gly Phe Lys Asp Asn Val Tyr Arg Gln Arg Arg Lys
                180                   185                   190
Tyr Phe Val Asp Val Ala Met Gly Tyr Lys Tyr Gly Gln Pro Ile Pro
                195                   200                   205
Arg Val Glu Tyr Thr Glu Glu Glu Thr Lys Thr Trp Gly Val Val Phe
    210                   215                   220
Arg Glu Leu Ser Lys Leu Tyr Pro Thr His Ala Cys Arg Glu Tyr Leu
225                   230                   235                   240
Lys Asn Phe Pro Leu Leu Thr Lys Tyr Cys Gly Tyr Arg Glu Asp Asn
                245                   250                   255
Val Pro Gln Leu Glu Asp Val Ser Met Phe Leu Lys Glu Arg Ser Gly
                260                   265                   270
Phe Thr Val Arg Pro Val Ala Gly Tyr Leu Ser Pro Arg Asp Phe Leu
                275                   280                   285
Ala Gly Leu Ala Tyr Arg Val Phe His Cys Thr Gln Tyr Ile Arg His
    290                   295                   300
Gly Ser Asp Pro Leu Tyr Thr Pro Glu Pro Asp Thr Cys His Glu Leu
305                   310                   315                   320
Leu Gly His Val Pro Leu Leu Ala Asp Pro Lys Phe Ala Gln Phe Ser
                325                   330                   335
Gln Glu Ile Gly Leu Ala Ser Leu Gly Ala Ser Asp Glu Asp Val Gln
                340                   345                   350
Lys Leu Ala Thr Cys Tyr Phe Phe Thr Ile Glu Phe Gly Leu Cys Lys
                355                   360                   365
Gln Glu Gly Gln Leu Arg Ala Tyr Gly Ala Gly Leu Leu Ser Ser Ile
    370                   375                   380
Gly Glu Leu Lys His Ala Leu Ser Asp Lys Ala Cys Val Lys Ala Phe
385                   390                   395                   400
Asp Pro Lys Thr Thr Cys Leu Gln Glu Cys Leu Ile Thr Thr Phe Gln
                405                   410                   415
Glu Ala Tyr Phe Val Ser Glu Ser Phe Glu Glu Ala Lys Glu Lys Met
                420                   425                   430
Arg Asp Phe Ala Lys Ser Ile Thr Arg Pro Phe Ser Val Tyr Phe Asn
                435                   440                   445
Pro Tyr Thr Gln Ser Ile Glu Ile Leu Lys Asp Thr Arg Ser Ile Glu
    450                   455                   460
Asn Val Val Gln Asp Leu Arg Ser Asp Leu Asn Thr Val Cys Asp Ala
465                   470                   475                   480
Leu Asn Lys Met Asn Gln Tyr Leu Gly Ile
                485                   490
```

```
<210> 3
<211> 1461
<212> DNA
<213> homo sapiens

<400> 3
atgatgatgt tttccagtaa atactgggca cggagagggt tttccctgga ttcagcagtg 60
cccgaagagc atcagctact tggcagctca acactaaata aacctaactc tggcaaaaat 120
```

```
gacgacaaag gcaacaaggg aagcagcaaa cgtgaagctg ctaccgaaag tggcaagaca 180
gcagttgttt tctccttgaa gaatgaagtt ggtggattgg taaaagcact gaggctcttt 240
caggaaaaac gtgtcaacat ggttcatatt gaatccagga aatctcggcg aagaagttct 300
gaggttgaaa tctttgtgga ctgtgagtgt gggaaaacag aattcaatga gctcattcag 360
ttgctgaaat ttcaaaccac tattgtgacg ctgaatcctc cagagaacat ttggacagag 420
gaagaagagc tagaggatgt gccctggttc cctcggaaga tctctgagtt agacaaatgc 480
tctcacagag ttctcatgta tggttctgag cttgatgctg accacccagg atttaaggac 540
aatgtctatc gacagagaag aaagtatttt gtggatgtgg ccatgggtta taaatatggt 600
cagcccattc cagggtggag tatactgaa gaagaaacta aaacttgggg tgttgtattc 660
cgggagctct ccaaactcta tcccactcat gcttgccgag agtatttgaa aaacttccct 720
ctgctgacta aatactgtgg ctacagagag gacaatgtgc ctcaactcga agatgtctcc 780
atgtttctga agaaaggtc tggcttcacg gtgaggccgg tggctggata cctgagccca 840
cgagactttc tggcaggact ggcctacaga gtgttccact gtacccagta catccggcat 900
ggctcagatc ccctctacac cccagaacca gacacatgcc atgaactctt gggacatgtt 960
ccactacttg cggatcctaa gtttgctcag ttttcacaag aaataggtct ggcgtctctg 1020
ggagcatcag atgaagatgt tcagaaacta gccacgtgct atttcttcac aatcgagttt 1080
ggcctttgca agcaagaagg gcaactgcgg gcatatggag caggactcct ttcctccatt 1140
ggagaattaa agcacgccct ttctgacaag gcatgtgtga aagcctttga cccaaagaca 1200
acttgcttac aggaatgcct tatcaccacc ttccaggaag cctactttgt ttcagaaagt 1260
tttgaagaag ccaaagaaaa gatgagggac tttgcaaagt caattacccg tcccttctca 1320
gtatacttca atccctacac acagagtatt gaaattctga aagacaccag aagtattgaa 1380
aatgtggtgc aggaccttcg cagcgacttg aatacagtgt gtgatgcttt aaacaaaatg 1440
aaccaatatc tggggatttg a                                          1461
```

```
<210> 4
<211> 486
<212> PRT
<213> homo sapiens

<400> 4
Met Met Met Phe Ser Ser Lys Tyr Trp Ala Arg Arg Gly Phe Ser Leu
1               5                   10                  15
Asp Ser Ala Val Pro Glu Glu His Gln Leu Leu Gly Ser Ser Thr Leu
            20                  25                  30
Asn Lys Pro Asn Ser Gly Lys Asn Asp Asp Lys Gly Asn Lys Gly Ser
        35                  40                  45
Ser Lys Arg Glu Ala Ala Thr Glu Ser Gly Lys Thr Ala Val Val Phe
    50                  55                  60
Ser Leu Lys Asn Glu Val Gly Gly Leu Val Lys Ala Leu Arg Leu Phe
65                  70                  75                  80
Gln Glu Lys Arg Val Asn Met Val His Ile Glu Ser Arg Lys Ser Arg
                85                  90                  95
Arg Arg Ser Ser Glu Val Glu Ile Phe Val Asp Cys Glu Cys Gly Lys
            100                 105                 110
Thr Glu Phe Asn Glu Leu Ile Gln Leu Leu Lys Phe Gln Thr Thr Ile
        115                 120                 125
Val Thr Leu Asn Pro Pro Glu Asn Ile Trp Thr Glu Glu Glu Glu Leu
    130                 135                 140
Glu Asp Val Pro Trp Phe Pro Arg Lys Ile Ser Glu Leu Asp Lys Cys
145                 150                 155                 160
Ser His Arg Val Leu Met Tyr Gly Ser Glu Leu Asp Ala Asp His Pro
                165                 170                 175
Gly Phe Lys Asp Asn Val Tyr Arg Gln Arg Arg Lys Tyr Phe Val Asp
            180                 185                 190
Val Ala Met Gly Tyr Lys Tyr Gly Gln Pro Ile Pro Arg Val Glu Tyr
        195                 200                 205
Thr Glu Glu Glu Thr Lys Thr Trp Gly Val Val Phe Arg Glu Leu Ser
    210                 215                 220
Lys Leu Tyr Pro Thr His Ala Cys Arg Glu Tyr Leu Lys Asn Phe Pro
225                 230                 235                 240
Leu Leu Thr Lys Tyr Cys Gly Tyr Arg Glu Asp Asn Val Pro Gln Leu
                245                 250                 255
Glu Asp Val Ser Met Phe Leu Lys Glu Arg Ser Gly Phe Thr Val Arg
```

```
              260                    265                    270
Pro Val Ala Gly Tyr Leu Ser Pro Arg Asp Phe Leu Ala Gly Leu Ala
          275                    280                    285
Tyr Arg Val Phe His Cys Thr Gln Tyr Ile Arg His Gly Ser Asp Pro
      290                    295                    300
Leu Tyr Thr Pro Glu Pro Asp Thr Cys His Glu Leu Leu Gly His Val
305                    310                    315                    320
Pro Leu Leu Ala Asp Pro Lys Phe Ala Gln Phe Ser Gln Glu Ile Gly
              325                    330                    335
Leu Ala Ser Leu Gly Ala Ser Asp Glu Asp Val Gln Lys Leu Ala Thr
              340                    345                    350
Cys Tyr Phe Phe Thr Ile Glu Phe Gly Leu Cys Lys Gln Glu Gly Gln
          355                    360                    365
Leu Arg Ala Tyr Gly Ala Gly Leu Leu Ser Ser Ile Gly Glu Leu Lys
      370                    375                    380
His Ala Leu Ser Asp Lys Ala Cys Val Lys Ala Phe Asp Pro Lys Thr
385                    390                    395                    400
Thr Cys Leu Gln Glu Cys Leu Ile Thr Thr Phe Gln Glu Ala Tyr Phe
              405                    410                    415
Val Ser Glu Ser Phe Glu Glu Ala Lys Glu Lys Met Arg Asp Phe Ala
              420                    425                    430
Lys Ser Ile Thr Arg Pro Phe Ser Val Tyr Phe Asn Pro Tyr Thr Gln
          435                    440                    445
Ser Ile Glu Ile Leu Lys Asp Thr Arg Ser Ile Glu Asn Val Val Gln
      450                    455                    460
Asp Leu Arg Ser Asp Leu Asn Thr Val Cys Asp Ala Leu Asn Lys Met
465                    470                    475                    480
Asn Gln Tyr Leu Gly Ile
              485
```

<210> 5
<211> 1458
<212> DNA
<213> homo sapiens

<400> 5

```
atgatgtttt ccagtaaata ctgggcacgg agagggtttt ccctggattc agcagtgccc 60
gaagagcatc agctacttgg cagctcaaca ctaaataaac ctaactctgg caaaaatgac 120
gacaaaggca caagggaag cagcaaacgt gaagctgcta ccgaaagtgg caagacagca 180
gttgtttttct ccttgaagaa tgaagttggt ggattggtaa aagcactgag gctctttcag 240
gaaaaacgtg tcaacatggt tcatattgaa tccaggaaat ctcggcgaag aagttctgag 300
gttgaaatct ttgtggactg tgagtgtggg aaaacagaat tcaatgagct cattcagttg 360
ctgaaattttc aaaccactat tgtgacgctg aatcctccag agaacatttg gacagaggaa 420
gaagagctag aggatgtgcc ctggttccct cggaagatct ctgagttaga caaatgctct 480
cacagagttc tcatgtatgg ttctgagctt gatgctgacc acccaggatt taaggacaat 540
gtctatcgac agagaagaaa gtattttgtg gatgtggcca tgggttataa atatggtcag 600
cccattccca gggtggagta tactgaagaa gaaactaaaa cttggggtgt tgtattccgg 660
gagctctcca aactctatcc cactcatgct tgccgagagt atttgaaaaa cttccctctg 720
ctgactaaat actgtggcta cagagaggac aatgtgcctc aactcgaaga tgtctccatg 780
tttctgaaag aaaggtctgg cttcacggtg aggccggtgg ctggatacct gagcccacga 840
gactttctgg caggactggc ctacagagtg ttccactgta cccagtacat ccggcatggc 900
tcagatcccc tctacacccc agaaccagac acatgccatg aactcttggg acatgttcca 960
ctacttgcgg atcctaagtt tgctcagttt tcacaagaaa taggtctggc gtctctggga 1020
gcatcagatg aagatgttca gaaactagcc acgtgctatt tcttcacaat cgagtttggc 1080
ctttgcaagc aagaagggca actgcgggca tatggagcag gactccttc ctccattgga 1140
gaattaaagc acgccctttc tgacaaggca tgtgtgaaag cctttgaccc aaagacaact 1200
tgcttacagg aatgccttat caccaccttc caggaagcct actttgtttc agaaagtttt 1260
gaagaagcca agaaaagat gagggacttt gcaaagtcaa ttacccgtcc cttctcagta 1320
tacttcaatc cctacacaca gagtattgaa attctgaaag acaccagaag tattgaaaat 1380
gtggtgcagg accttcgcag cgacttgaat acagtgtgtg atgctttaaa caaaatgaac 1440
caatatctgg ggatttga                                                 1458
```
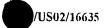
4/7

<210> 6
<211> 485
<212> PRT
<213> homo sapiens

<400> 6

```
Met Met Phe Ser Ser Lys Tyr Trp Ala Arg Arg Gly Phe Ser Leu Asp
1               5                   10                  15
Ser Ala Val Pro Glu Glu His Gln Leu Leu Gly Ser Ser Thr Leu Asn
                20                  25                  30
Lys Pro Asn Ser Gly Lys Asn Asp Asp Lys Gly Asn Lys Gly Ser Ser
            35                  40                  45
Lys Arg Glu Ala Ala Thr Glu Ser Gly Lys Thr Ala Val Val Phe Ser
        50                  55                  60
Leu Lys Asn Glu Val Gly Gly Leu Val Lys Ala Leu Arg Leu Phe Gln
65                  70                  75                  80
Glu Lys Arg Val Asn Met Val His Ile Glu Ser Arg Lys Ser Arg Arg
                85                  90                  95
Arg Ser Ser Glu Val Glu Ile Phe Val Asp Cys Glu Cys Gly Lys Thr
                100                 105                 110
Glu Phe Asn Glu Leu Ile Gln Leu Leu Lys Phe Gln Thr Thr Ile Val
            115                 120                 125
Thr Leu Asn Pro Pro Glu Asn Ile Trp Thr Glu Glu Glu Glu Leu Glu
        130                 135                 140
Asp Val Pro Trp Phe Pro Arg Lys Ile Ser Glu Leu Asp Lys Cys Ser
145                 150                 155                 160
His Arg Val Leu Met Tyr Gly Ser Glu Leu Asp Ala Asp His Pro Gly
                165                 170                 175
Phe Lys Asp Asn Val Tyr Arg Gln Arg Arg Lys Tyr Phe Val Asp Val
            180                 185                 190
Ala Met Gly Tyr Lys Tyr Gly Gln Pro Ile Pro Arg Val Glu Tyr Thr
        195                 200                 205
Glu Glu Glu Thr Lys Thr Trp Gly Val Val Phe Arg Glu Leu Ser Lys
    210                 215                 220
Leu Tyr Pro Thr His Ala Cys Arg Glu Tyr Leu Lys Asn Phe Pro Leu
225                 230                 235                 240
Leu Thr Lys Tyr Cys Gly Tyr Arg Glu Asp Asn Val Pro Gln Leu Glu
                245                 250                 255
Asp Val Ser Met Phe Leu Lys Glu Arg Ser Gly Phe Thr Val Arg Pro
                260                 265                 270
Val Ala Gly Tyr Leu Ser Pro Arg Asp Phe Leu Ala Gly Leu Ala Tyr
        275                 280                 285
Arg Val Phe His Cys Thr Gln Tyr Ile Arg His Gly Ser Asp Pro Leu
    290                 295                 300
Tyr Thr Pro Glu Pro Asp Thr Cys His Glu Leu Leu Gly His Val Pro
305                 310                 315                 320
Leu Leu Ala Asp Pro Lys Phe Ala Gln Phe Ser Gln Glu Ile Gly Leu
                325                 330                 335
Ala Ser Leu Gly Ala Ser Asp Glu Asp Val Gln Lys Leu Ala Thr Cys
        340                 345                 350
Tyr Phe Phe Thr Ile Glu Phe Gly Leu Cys Lys Gln Glu Gly Gln Leu
    355                 360                 365
Arg Ala Tyr Gly Ala Gly Leu Leu Ser Ser Ile Gly Glu Leu Lys His
    370                 375                 380
Ala Leu Ser Asp Lys Ala Cys Val Lys Ala Phe Asp Pro Lys Thr Thr
385                 390                 395                 400
Cys Leu Gln Glu Cys Leu Ile Thr Thr Phe Gln Glu Ala Tyr Phe Val
                405                 410                 415
Ser Glu Ser Phe Glu Glu Ala Lys Glu Lys Met Arg Asp Phe Ala Lys
                420                 425                 430
Ser Ile Thr Arg Pro Phe Ser Val Tyr Phe Asn Pro Tyr Thr Gln Ser
            435                 440                 445
Ile Glu Ile Leu Lys Asp Thr Arg Ser Ile Glu Asn Val Val Gln Asp
```

```
        450                    455                    460
Leu Arg Ser Asp Leu Asn Thr Val Cys Asp Ala Leu Asn Lys Met Asn
465                    470                    475                    480
Gln Tyr Leu Gly Ile
                485
```

<210> 7
<211> 1455
<212> DNA
<213> homo sapiens

<400> 7
```
atgttttcca gtaaatactg ggcacggaga gggtttttccc tggattcagc agtgcccgaa 60
gagcatcagc tacttggcag ctcaacacta aataaaccta actctggcaa aaatgacgac 120
aaaggcaaca agggaagcag caaacgtgaa gctgctaccg aaagtggcaa gacagcagtt 180
gttttctcct tgaagaatga agttggtgga ttggtaaaag cactgaggct ctttcaggaa 240
aaacgtgtca acatggttca tattgaatcc aggaaatctc ggcgaagaag ttctgaggtt 300
gaaatctttg tggactgtga gtgtgggaaa acagaattca atgagctcat tcagttgctg 360
aaatttcaaa ccactattgt gacgctgaat cctccagaga acatttggac agaggaagaa 420
gagctagagg atgtgccctg gttccctcgg aagatctctg agttagacaa atgctctcac 480
agagttctca tgtatggttc tgagcttgat gctgaccacc caggatttaa ggacaatgtc 540
tatcgacaga gaagaaagta tttgtggat gtggccatgg gttataaata tggtcagccc 600
attcccaggg tggagtatac tgaagaagaa actaaaactt ggggtgttgt attccgggag 660
ctctccaaac tctatcccac tcatgcttgc cgagagtatt tgaaaaactt ccctctgctg 720
actaaatact gtggctacag agaggacaat gtgcctcaac tcgaagatgt ctccatgttt 780
ctgaaagaaa ggtctggctt cacggtgagg ccggtggctg atacctgag cccacgagac 840
tttctggcag gactggccta cagagtgttc cactgtaccc agtacatccg gcatggctca 900
gatcccctct acacccccaga accagacaca tgccatgaac tcttgggaca tgttccacta 960
cttgcggatc ctaagtttgc tcagttttca caagaaatag gtctggcgtc tctgggagca 1020
tcagatgaag atgttcagaa actagccacg tgctatttct tcacaatcga gtttggcctt 1080
tgcaagcaag aagggcaact gcgggcatat ggagcaggac tccttcctc cattggagaa 1140
ttaaagcacg ccctttctga caaggcatgt gtgaaagcct tgacccaaa gacaacttgc 1200
ttacaggaat gccttatcac caccttccag gaagcctact ttgtttcaga aagtttgaa 1260
gaagccaaag aaaagatgag ggactttgca aagtcaatta cccgtccctt ctcagtatac 1320
ttcaatccct acacacagag tattgaaatt ctgaaagaca ccagaagtat tgaaaatgtg 1380
gtgcaggacc ttcgcagcga cttgaataca gtgtgtgatg cttttaaacaa aatgaaccaa 1440
tatctgggga tttga                                                   1455
```

<210> 8
<211> 484
<212> PRT
<213> homo sapiens

<400> 8
```
Met Phe Ser Ser Lys Tyr Trp Ala Arg Arg Gly Phe Ser Leu Asp Ser
  1               5                  10                  15
Ala Val Pro Glu Glu His Gln Leu Leu Gly Ser Ser Thr Leu Asn Lys
               20                  25                  30
Pro Asn Ser Gly Lys Asn Asp Asp Lys Gly Asn Lys Gly Ser Ser Lys
           35                  40                  45
Arg Glu Ala Ala Thr Glu Ser Gly Lys Thr Ala Val Val Phe Ser Leu
       50                  55                  60
Lys Asn Glu Val Gly Gly Leu Val Lys Ala Leu Arg Leu Phe Gln Glu
65                  70                  75                  80
Lys Arg Val Asn Met Val His Ile Glu Ser Arg Lys Ser Arg Arg Arg
               85                  90                  95
Ser Ser Glu Val Glu Ile Phe Val Asp Cys Glu Cys Gly Lys Thr Glu
              100                 105                 110
Phe Asn Glu Leu Ile Gln Leu Leu Lys Phe Gln Thr Thr Ile Val Thr
           115                 120                 125
Leu Asn Pro Pro Glu Asn Ile Trp Thr Glu Glu Glu Leu Glu Asp
           130                 135                 140
```

```
Val Pro Trp Phe Pro Arg Lys Ile Ser Glu Leu Asp Lys Cys Ser His
145                 150                 155                 160
Arg Val Leu Met Tyr Gly Ser Glu Leu Asp Ala Asp His Pro Gly Phe
                165                 170                 175
Lys Asp Asn Val Tyr Arg Gln Arg Arg Lys Tyr Phe Val Asp Val Ala
            180                 185                 190
Met Gly Tyr Lys Tyr Gly Gln Pro Ile Pro Arg Val Glu Tyr Thr Glu
        195                 200                 205
Glu Glu Thr Lys Thr Trp Gly Val Val Phe Arg Glu Leu Ser Lys Leu
    210                 215                 220
Tyr Pro Thr His Ala Cys Arg Glu Tyr Leu Lys Asn Phe Pro Leu Leu
225                 230                 235                 240
Thr Lys Tyr Cys Gly Tyr Arg Glu Asp Asn Val Pro Gln Leu Glu Asp
            245                 250                 255
Val Ser Met Phe Leu Lys Glu Arg Ser Gly Phe Thr Val Arg Pro Val
        260                 265                 270
Ala Gly Tyr Leu Ser Pro Arg Asp Phe Leu Ala Gly Leu Ala Tyr Arg
        275                 280                 285
Val Phe His Cys Thr Gln Tyr Ile Arg His Gly Ser Asp Pro Leu Tyr
    290                 295                 300
Thr Pro Glu Pro Asp Thr Cys His Glu Leu Leu Gly His Val Pro Leu
305                 310                 315                 320
Leu Ala Asp Pro Lys Phe Ala Gln Phe Ser Gln Glu Ile Gly Leu Ala
            325                 330                 335
Ser Leu Gly Ala Ser Asp Glu Asp Val Gln Lys Leu Ala Thr Cys Tyr
            340                 345                 350
Phe Phe Thr Ile Glu Phe Gly Leu Cys Lys Gln Glu Gly Gln Leu Arg
            355                 360                 365
Ala Tyr Gly Ala Gly Leu Leu Ser Ser Ile Gly Glu Leu Lys His Ala
    370                 375                 380
Leu Ser Asp Lys Ala Cys Val Lys Ala Phe Asp Pro Lys Thr Thr Cys
385                 390                 395                 400
Leu Gln Glu Cys Leu Ile Thr Thr Phe Gln Glu Ala Tyr Phe Val Ser
            405                 410                 415
Glu Ser Phe Glu Glu Ala Lys Glu Lys Met Arg Asp Phe Ala Lys Ser
            420                 425                 430
Ile Thr Arg Pro Phe Ser Val Tyr Phe Asn Pro Tyr Thr Gln Ser Ile
        435                 440                 445
Glu Ile Leu Lys Asp Thr Arg Ser Ile Glu Asn Val Val Gln Asp Leu
    450                 455                 460
Arg Ser Asp Leu Asn Thr Val Cys Asp Ala Leu Asn Lys Met Asn Gln
465                 470                 475                 480
Tyr Leu Gly Ile
```

THIS PAGE BLANK (USPTO)